

VIDEO TRANSCODER BASED JOINT VIDEO AND STILL IMAGE PIPELINE WITH STILL BURST MODE

Technical Field

The technical field relates to video imaging systems, and, in particular, to joint video and still image pipelines.

Background

Digital cameras are widely used to acquire high resolution still image photographs. Digital video cameras are also used to record home videos, television programs, movies, concerts, or sports events on a magnetic disk or optical DVD for storage or transmission through communications channels. Some commercial cameras are able to take both digital video and digital still image photographs. However, most of these cameras required a user to switch between a video recording mode and a digital still image mode. Separate pipelines are generally used for each of the video recording and still image modes. Examples of these cameras include SANYO ID-SHOT® and CANNON POWERSHOT S300®. The SANYO ID-SHOT® uses an optical disk, whereas the CANNON POWERSHOT S300® uses synchronous dynamic random access memory (SDRAM). However, both cameras are still image cameras that have the capability of taking video clips, using separate pipelines.

Other cameras use a single software pipeline to acquire both digital video and low quality still images by taking one of the video frames as is, and storing the particular video frame as a high resolution still image. Examples of such cameras include JVC GR-DVL9800®, which is a digital video camera that allows a user to take a picture at certain point in time. However, the pictures taken generally are of low quality, because a low resolution video pipeline is used to generate the high resolution still image pictures.

When still images are acquired in burst mode, current cameras try to process both pipelines independently. If a single hardware processing pipeline is used, a large frame buffer may be needed to store video frames while the burst mode still images are processed. However, a large frame buffer is costly, and build up delay on the video side may be undesirable.

Summary

A method and corresponding apparatus for concurrently processing digital video frames and high resolution still images in burst mode includes acquiring regular size video frames and high resolution still image frames in burst mode from one or more

image sensors, and downsampling the regular size video frames into reduced size video frames in real time so that the reduced size video frames have frame sizes smaller than the regular size video frames. The method further includes processing the high resolution still image frames acquired during the burst mode using a high resolution still image pipeline, and processing the reduced size frames using a video pipeline. The high resolution still image frames are processed concurrently with the reduced size video frames.

In an embodiment of the method, the reduced size video frames are upsampled using motion estimation and information from the high resolution still image frames. In another embodiment of the method, the high resolution still image frames are downsampled into downsampled still image frames that have the same frame size as the upsampled video frames. Blocks in the downsampled still image frames form a block pool. In yet another embodiment, blocks in the block pool are compared with corresponding blocks in the upsampled video frames until a best match block is found. The best match block is then copied into the corresponding blocks in the upsampled video frames to compensate the loss in video quality.

Description of the Drawings

The preferred embodiments of the method and apparatus for concurrently processing digital video frames and high resolution still images in burst mode will be described in detail with reference to the following figures, in which like numerals refer to like elements, and wherein:

Figure 1 illustrates an exemplary operation of an exemplary joint video and still image pipeline;

Figure 2 illustrates a preferred embodiment of a video camera system using the exemplary joint video and still image pipeline of Figure 1;

Figure 3 illustrates an exemplary hardware implementation of the exemplary joint video and still image pipeline of Figure 1;

Figures 4A- 4C are flow charts describing in general the exemplary joint video and still image pipeline of Figure 1;

Figures 5A-5E illustrate an exemplary motion estimation technology used by the video camera system of Figure 2;

Figure 6 is a flow chart illustrating an exemplary method for concurrently processing digital video frames and high resolution still images in burst mode; and

Figure 7 is a flow chart illustrating the exemplary motion estimation technology shown in Figures 5A-5E.

Detailed Description

A digital video camera system may utilize a joint video and still image pipeline that simultaneously acquires, processes, transmits and/or stores digital video and high resolution digital still image photographs. The joint pipeline may include a video pipeline optimized for digital video frames and a high resolution still image pipeline optimized for high resolution digital still images. The digital video camera system may also concurrently acquire and process video frames and high resolution still image in burst mode by downsampling the video frames into reduced size video frames so that the single processing pipeline can allot more time to process the high resolution still images in real time. The loss in video quality may be compensated using motion estimation and information from the high resolution still images.

Figure 1 illustrates an exemplary operation of an exemplary joint video and still image pipeline, which is capable of simultaneously capturing digital video frames 120 and high resolution digital still image frames 110. The video frames 120 may be acquired at, for example, 30 frames per second (fps). During video frame acquisition, a snapshot 102 may be taken to acquire a particular still image frame 110 in high resolution, which is then processed. During the high resolution still image processing, all incoming video frames 120 that are captured during that time may be temporarily stored, i.e., buffered, in a frame buffer 330 (shown in Figure 3) before being processed. Both the video frames 120 and the high resolution still image frame 110 may be stored or transmitted through communications channels, such as a network.

Figure 2 illustrates a preferred embodiment of a video camera system 200 using the exemplary joint video and still image pipeline. In this embodiment, a video pipeline 220 and a high resolution still image pipeline 210 share a same high resolution image sensor 240. The high resolution image sensor 240, which may be a charge coupled device (CCD) sensor or a complimentary metal oxide semiconductor (CMOS) sensor, may take high resolution still image frames 110 while acquiring medium resolution video frames 120. This embodiment is inexpensive because the video camera system 200 uses one hardware processing pipeline 300 (shown in Figure 3) with one image sensor 240 and one processor 360 (shown in Figure 3).

The image sensor 240 typically continuously acquires high resolution video frames 120 at a rate of, for example, 30 fps. Each of the high resolution video frames 120

1 may be converted into a high resolution still image photograph 110. When a user is not
2 interested in taking a high resolution still image photograph 110, the only pipeline
3 running may be the video pipeline 220, which acquires high resolution video frames 120,
4 and downsamples the frames to medium resolution (for example, 640x480), then
5 processes the medium resolution video frames 120. When the user wants to acquire a
6 high resolution still image frame 110, the image acquired by the high resolution image
7 sensor 240 can be used both in the video pipeline 220 as well as in the high resolution still
8 image pipeline 210 (described in detail later).

9 The video camera system 200 may include a storage device 250 and a connection
10 with a communications channel/network 260, such as the Internet or other type of
11 computer or telephone networks. The storage device 250 may include a hard disk drive,
12 floppy disk drive, CD-ROM drive, or other types of non-volatile data storage, and may
13 correspond with various databases or other resources. After the video frames 120 and the
14 high resolution still image frames 110 are acquired, the video frames 120 and the high
15 resolution still image frames 110 may be stored in the storage device 250 or transmitted
16 through the communication channel 260. The video camera system 200 may also include
17 a video transcoding agent 270 for encoding the video frames 120 into a standard format,
18 for example, motion picture expert group-2 (MPEG-2).

19 Figure 3 illustrates an exemplary hardware implementation of the preferred
20 embodiment of the exemplary joint video and still image pipeline. This embodiment
21 includes the single hardware processing pipeline 300 supporting two software pipelines.
22 A sensor controller 310 may be controlled by a user to retrieve high resolution mosaiced
23 still image frames 110 at a rate of, for example, one every thirtieth of a second to generate
24 a video signal. The sensor controller 310 may then store the selected high resolution still
25 image frames 110 into a memory 320. The memory 320 may include random access
26 memory (RAM) or similar types of memory. Next, the high resolution still image frames
27 110 may be processed using the processor 360, which may be a microprocessor 362, an
28 ASIC 364, or a digital signal processor 366. The ASIC 362 performs algorithms quickly,
29 but is typically application specific and only performs a specific algorithm. On the other
30 hand, the microprocessor 362 or the digital signal processor 366 may perform many other
31 tasks. The processor 360 may execute information stored in the memory 320 or the
32 storage device 250, or information received from the Internet or other network 260. The
33 digital video and still image data may be copied to various components of the pipeline
34 300 over a data bus 370.

1 In the video pipeline 220, the processor 360 may downsample, demosaic, and
2 color correct the video frames 120. Next, the processor 360 may compress and transmit
3 the video frames 120 through an input/output (I/O) unit 340. Alternatively, the video
4 frames 120 may be stored in the storage device 250.

5 Both pipelines may be executed concurrently, i.e., acquiring high resolution still
6 image photographs 110 during video recording. A frame buffer 330 may store video
7 frames 120 while the processor 360 is processing the high resolution still image frame
8 110. The sensor controller 310 may still capture video frames 120 at a rate of, for
9 example, 30 fps, and store the video frames 120 into the memory 320. The processor 360
10 may downsample the video frames 120 and send the downsampled video frames 120 into
11 the frame buffer 330. The frame buffer 330 may store the downsampled video frames
12 120 temporarily without further processing. This may incur some delay in the video
13 pipeline 220 if the video is directly transmitted through the communications channel 260.
14 However, this delay may be compensated by a similar buffer on the receiver end. During
15 video frame buffering, the high resolution still image frame 110 may be processed by the
16 processor 360, using complex algorithms. At the same time, the video frames 120 are
17 continuously stored into the memory 320, downsampled, and sent into the frame buffer
18 330 to be stored. Minimal video buffering is preferred and may be achieved through real
19 time video downsampling (described in detail later).

20 Although the video camera system 200 is shown with various components, one
21 skilled in the art will appreciate that the video camera system 200 can contain additional
22 or different components. In addition, although the video frames 120 and the still image
23 frames 110 are described as being stored in memory, one skilled in the art will appreciate
24 that the video frames 120 and the still image frames 110 can also be stored on or read
25 from other types of computer program products or computer-readable media, such as
26 secondary storage devices, including hard disks, floppy disks, or CD-ROM; a carrier
27 wave from the Internet or other network; or other forms of RAM or ROM. The
28 computer-readable media may include instructions for controlling the video camera
29 system 200 to perform a particular method.

30 Figures 4A- 4C are flow charts describing in general operation of the exemplary
31 joint video and still image pipeline. Referring to Figure 4A, operation of the video
32 pipeline 220, shown on the left, typically results in continuous processing of video frames
33 120. Operation of the high resolution still image pipeline 210, shown on the right,

1 typically results in processing a high resolution still image frame 110 every time the user
2 wants to acquire a high resolution photograph 110.

3 After raw pixel video data of video frames 120 are acquired, for example, at
4 1024x1008 and 30 fps (block 400), the video frames 120 may be downsampled and
5 demosaiced in order to save memory space (block 410). Then, the frame buffer 330 may
6 buffer the video frames 120 while the high resolution still image frame 110 is being
7 acquired, processed, stored, and/or transmitted (block 420). Alternatively, demosaicing
8 may be performed after the video frames 120 are buffered. Thereafter, the video pipeline
9 220 may start emptying the frame buffer 330 as fast as possible, and performing color
10 correction, compression, storage and/or transmission (blocks 430, 440, 450). Once the
11 frame buffer 330 is emptied, another high resolution still image frame 110 may be
12 acquired.

13 For high resolution still image frames 110, sophisticated demosaicing may be
14 performed (block 412), followed by high quality color correction (block 432). The high
15 resolution still image frames 110 may optionally be compressed (block 442), and then
16 stored and/or transmitted through similar communications channels 260 (block 452).

17 Figure 4B illustrates in detail the operation of the high resolution still image
18 pipeline 210. The sophisticated demosaicing process (block 412) utilizes a high quality
19 demosaicing algorithm that generates a high quality color image from the originally
20 mosaiced image acquired by the image sensor 240. The demosaicing process is a time
21 consuming filtering operation, which may gamma-correct the input if the image sensor
22 240 has not done so, resulting in excellent color image quality with almost no
23 demosaicing artifacts. For example, demosaicing for high resolution still image frames
24 110 may filter the original image with a 10x10 linear filter. The algorithm takes into
25 account the lens used for acquisition, as well as the spectral sensitivity of each of the
26 color filters on the mosaic.

27 Once the high resolution still image frame 110 is demosaiced, the high resolution
28 still image frame 110 may be color corrected depending on the illumination present at the
29 time of the capture (block 432). Complex transformation matrixes may be involved to
30 restore accurate color to the high resolution still image frames 110, in order to generate an
31 excellent photograph. The color correction algorithms, may be similar to the algorithm
32 used in the HP-PHOTOSMART 618®.

Figure 4C illustrates in detail the operation of the video pipeline 220. A high quality video pipeline 220 may be very demanding in terms of computation. Because the video processing needs to be achieved at, for example, 30 fps, downsampling may be fast. In addition, lower resolution video frames 120 (for example, 640x480 pixels) demands much less quality demosaicing (block 410), because the human visual system may not notice certain artifacts at high video frame rates. For example, demosaicing for video frames 120 may filter the original image with a 4x4 linear filter. Similarly, color correction may be simpler because high quality is not needed on the video side (block 430).

When a user acquires high resolution still image frames 110 in burst mode, the digital video camera system 200 downsamples the video frames 120 into reduced size (downsampled) video frames. For example, the size of video frames may be downsampled from 640x480 pixels to 320x240 pixels. With lower quality video processing, the joint video and still image pipeline can spend more time processing the high resolution still image frames 110. Therefore, the video camera system 200 leverages between the video pipeline 220 and the high resolution still image pipeline 210 to reduce the computational burden, so that the burst mode high resolution still image frames 110 can be processed in real time with high priority, concurrently with the reduced size video frame processing.

After the high resolution still image frames 110 are demosaiced, color corrected, and compressed (with or without loss), the high resolution still image frames 110 may be used to reset MPEG encoding process as intraframes (I-frames). I-frames are frames not compressed depending on previous or future frames, i.e., stand alone compressed frames. Accordingly, all compression algorithms need to start with an I-frame, and all other frames may be compressed based on the I-frame.

The reduced size (downsampled) video frames 120 can be stored with near lossless compression or no compression at all, in order to recover as good a quality as possible. Finally the reduced size video frames 120 are upsampled to the original video size (for example, 640x480 pixels) using, for example, motion estimation. Information from the high resolution still image frames 110 may be used to recover the loss in quality of the video frames 120. Since many high resolution still image frames 110 are taken in burst mode, video frame quality can be improved through motion compensation and motion estimation technology, using information from the high resolution still image frames 110.

1 Motion estimation is described, for example, in "Image and Video Compression
2 Standards" by Bhaskaran and Konstantinides, chapters 4.4 – 4.7, which is incorporated
3 herein by reference. In motion estimation, a block of image is typically compared with a
4 previous image frame to detect changes, then matched into the previous frame. For
5 example, if a person moves her head and mouth with the background image unchanged,
6 the background does not need to be compressed every time a new image is acquired. A
7 motion detection system detects that the background is not moving, and copies the
8 background from the previous frame. Motion estimation technology estimates changes in
9 image frames and predicts future frames based on the changes. If, for example, one high
10 resolution still image frame 110 is taken every eight video frames 120, the high resolution
11 still image frame 110 can be downsampled and matched into other video frames 120 to
12 compensate the loss in quality of the video frames 120 (described in detail with respect to
13 Figures 5A-5C).

14 Within the video camera system 200, the reduced size video frames 120 may be
15 kept in a nonstandard proprietary format. The video transcoding agent 270, which
16 typically runs on the video camera system 200, transcodes the proprietary loss-less (or
17 near loss-less) compressed video frames 120 into a standard format, for example, MPEG-
18 2. Alternatively, the video transcoding agent 270 may run on a docking station or on the
19 host personal computer (PC).

20 For example, high resolution still image frames 110 are typically 2 mega (M)
21 pixel in size, whereas the size of a video frame is typically $640 \times 480 = 307,200$ pixels. If
22 the video frames 120 are acquired at 30 fps, the video pixel size is 30 x 307,200 pixels per
23 second (pps). Therefore, the size of one high resolution still image is approximately 6.41
24 times the size of a video frame ($2,000,000/307,200 = 6.41$). Processing rate per pixel
25 typically needs to be maintained approximately constant with and without burst mode,
26 given maximum processing power of the pipelines.

27 In the following exemplary equations (for illustration purposes only), VS stands
28 for video frame size; SS stands for still image size, which equals $6.41 \times VS$; RF stands for
29 reduction factor, which is the number of times the size of video frames 120 needs to be
30 reduced in order to take a high resolution still image frame 110 simultaneously; and RS
31 stands for reduced size, which equals VS/RF . The following equations calculates N,
32 which stands for the number of video frames 120 between two high resolution still image
33 frames 110. In other words, one high resolution still image frame 110 can be acquired
34 and processed simultaneously with N video frames 120 using the limited hardware

1 processing power. The reduced quality of the video frames 120 can be compensated
2 during upsampling using motion estimation and information from the high resolution still
3 image frames 110. $N*VS = SS + (N-1) RS$

$$4 \quad N*VS = 6.41*VS + (N-1) VS/RF$$

$$5 \quad RF*N = 6.41*RF + N-1$$

6 For example, if $RF = 4$, $N = 8.21$. Therefore, if video size is reduced by a factor
7 of four, one high resolution still image frame 110 can be acquired every 8.21 video
8 frames 120.

9 If, for example, $RF = 2$ instead, N becomes 11.82. Therefore, if video size is
10 reduced by a factor of two, one high resolution still image frame 110 can be acquired
11 every 11.82 video frames. In other words, a user needs to wait longer to process one high
12 resolution still image frame 110.

13 Finally, if $RF = 1$, the entire video frames 120 may be stored without any
14 compression, requiring extra processing power and occupying extra storage. Since no
15 video size reduction occurs, no extra time is available to process the high resolution still
16 image frames 110. Therefore, with limited hardware processing power, the video frames
17 120 and the burst mode high resolution still image frames 110 cannot be processed
18 simultaneously.

19 Accordingly, as N increases, video quality worsens, but more burst mode still
20 image frames 110 may be acquired per second. The total power of the hardware
21 processing pipeline is split between a reduced size video pipeline 220 and the high
22 resolution still image pipeline 210.

23 Figures 5A-5E illustrate an exemplary motion estimation technology used by the
24 video camera system 200 of Figure 2. Referring to Figure 5A, one high resolution still
25 image frame 110 is taken every four video frames 120. Frames #1 and #5, 111 and 112,
26 respectively, are high resolution still image frames. Frames #2, #3, and #4, 121, 122,
27 123, respectively, are reduced size (downsampled) video frames. Referring to Figure 5B,
28 in this example, the high resolution still image frame 111 (frame #1) is shown with one or
29 more 32x32 blocks, the reduced size video frame 121 (frame #2) is shown with one or
30 more 4x4 blocks, and a regular size video frame (not shown) typically has 8x8 blocks.

31 Referring to Figure 5C, the processors 360 upsample the 4x4 block in the reduced
32 size video frame 121 (frame #2) into an 8x8 block in an upsampled video frame 221. The
33 upsampled video frame 221 typically has the same frame size as the regular size video
34 frame. Likewise, all of the 4x4 blocks in the reduced size video frame 121 are upsampled

1 into 8x8 blocks. The upsampling method may be simple or sophisticated, using, for
2 example, bilinear interpolation, bicubic interpolation, or resolution synthesis.

3 Referring to Figure 5D, the high resolution still image frame 111 (frame #1) is
4 downsampled into a downsampled still image frame 511 (shown in Figure 5E), which
5 becomes an I-frame. The downsampled still image frame 511 typically has the same
6 frame size as the upsampled video frame 221. Next, all of the 32x32 blocks in the high
7 resolution still image frame 111 are downsampled into multiple 8x8 video frame size
8 blocks, forming a block pool 550. All possible 32x32 blocks are considered in order to
9 maximize quality. Then, all of the 8x8 blocks in the block pool 550 are compared with
10 the 8x8 block in the upsampled video frame 221 until a best match block is found, using
11 various motion estimation algorithms. Finally, the selected best match 8x8 block in the
12 block pool 550 is copied into the corresponding position in the upsampled video frame
13 221. Similarly process needs to be performed for all of the 8x8 blocks in upsampled
14 video frame 221, generating a new I-frame 521 for frame #2.

15 Referring to Figure 5E, frames #3 and #4 can be predicted from frame #2 in order
16 to increase compression ratio. For example, the processors 360 encode motion vectors of
17 frame #2 (with video resolution of 640x480 pixels) in frame #3, which is upsampled from
18 4x4 block size to 8x8 block, generating frame 522. Error is encoded if necessary. Frame
19 #4 523 can be generated in a similar fashion. In this example, two predicted (P) frames
20 522, 523 follow two I-frames 511, 521.

21 Figure 6 is a flow chart illustrating the method for concurrently processing digital
22 video frames and high resolution still image frames in burst mode. The video camera
23 system 200 acquires video frames 120 and high resolution still image frames 110 in burst
24 mode (block 610). The processors 360 then downsample the video frames 120 into
25 reduced size video frames so that the single processing pipeline can allot more time to
26 process the high resolution still image frames 110 (block 620). Next, the processors 360
27 process the high resolution still image frames 110 in real time (block 630), concurrently
28 with the reduced size video frames (block 640). If extra processing power is available,
29 the processors 360 upsample the reduced size video frames into regular size video frames
30 using motion estimation technology and information from the high resolution still image
31 frames acquired in burst mode (block 650). Alternatively, the upsampling and encoding
32 process may be achieved during transmission/download time with the video transcoding
33 agent 270. In other words, if the video frames are not fully encoded due to lack of

1 memory space or processing power, the video frames can be encoded fully at download
2 time by the video transcoding agent 270.

3 Figure 7 is a flow chart illustrating the motion estimation technology used by the
4 video camera system 200. First, the processors 360 upsample the reduced size video
5 frame 121, so that the frame sizes of the upsampled video frame 221 are the same as the
6 regular size video frames originally acquired by the image sensor 240 (block 710). Next,
7 the processors 360 downsample the high resolution still image frame 111 into a
8 downsampled still image frame 511, which has the same frame size as the upsampled
9 video frame 221. All blocks in the high resolution still image frame 111 are
10 downsampled into multiple video frame size blocks, forming a block pool 550 (block
11 720). Then, for each block in the upsampled video frame 221, the processors 360
12 compare all of the blocks in the block pool 550 with the corresponding block in the
13 upsampled video frame 221, until a best match block is found (block 730). Finally, the
14 best match block in the block pool 550 is copied into the corresponding block in the
15 upsampled video frame 221, gradually generating a new I-frame 521 after similar process
16 is performed for all of the blocks in the upsampled video frame 221 (block 740).

17 While the method and apparatus for concurrently processing digital video frames
18 and high resolution still images in burst mode have been described in connection with an
19 exemplary embodiment, those skilled in the art will understand that many modifications
20 in light of these teachings are possible, and this application is intended to cover any
21 variations thereof.